# Second exam January 10, 2007
# Statistics II

1. (6 points) Uma determinada universidade tem duas faculdades de gestão, uma localizada na cidade $A$ e outra localizada na cidade $B$. Na faculdade localizada na cidade $A$, a pessoa que trata das admissões a um determinado programa de MBA verificou que, historicamente, as notas finais de curso dos candidatos seguem uma distribuição normal com desvio padrão igual a 0.45. Foi recolhida uma amostra aleatória de 25 candidaturas para o corrente ano, verificando-se que a média das notas finais de curso foi de 2.9 (numa escala de 0 a 5).

   (a) Encontre um intervalo de confiança a 95% para a nota média.

   **Solution:** Seja $X_A$: notas finais de curso dos candidatos a $A \Longrightarrow X_A \sim \mathcal{N}\left(\mu_A, \sigma_A^2 = 0.45^2\right)$

   Sabemos que
   $$Z = \frac{\overline{X}_A - \mu_A}{\sigma_A / N} \sim \mathcal{N}(0,1)$$
   O nível de confiança é $1-\alpha = 0.95 \iff \alpha = 0.05$. Assim, $z_{0.025} = 1.96$. Então, o intervalo de confiança a 95% para a nota média é
   $$[\mu_A]_{95\%} = \left[\overline{X}_A \pm z_{\alpha/2}\frac{\sigma_A}{\sqrt{N}}\right] = \left[2.9 \pm 1.96\,\frac{0.45}{\sqrt{25}}\right] = [2.7236; 3.076]$$

   (b) Com base nestes resultados, um professor de estatística calculou um intervalo de confiança para a nota média entre 2.81 e 2.99. Qual a percentagem de intervalos assim calculados que contêm a verdadeira nota média da população?

   **Solution:** Agora temos que calcular o valor de $\alpha$. Sabemos que
   $$\mathbb{P}\left(\overline{X}_A < z_{\alpha/2}\frac{\sigma_A}{\sqrt{N}} < \mu_A < \overline{X}_A + z_{\alpha/2}\frac{\sigma_A}{\sqrt{N}}\right) = 1 - \alpha$$
   A amplitude do intervalo é
   $$\Delta = 2 \cdot z_{\alpha/2}\frac{\sigma_A}{\sqrt{N}} = 2.99 - 2.81 = 0.18$$
   No exercício, $2 \cdot z_{\alpha/2}\dfrac{0.45}{5} = 0.18 \iff z_{\alpha/2} = 1$. Usando a tabela da distribuição normal obtém-se que $\alpha/2 = 1 - 0.8413 \iff \alpha = 0.3174 \iff 1-\alpha = 0.6826$. A percentagem de intervalos assim calculados que contêm o verdadeiro valor de $\mu_A$ é de 68.26%.

   (c) Qual o número de candidaturas que é necessário incluir na amostra para que seja possível obter um intervalo de confiança a 95% para a nota média com uma margem de erro de 0.05?

**Solution:** Agora a margem de erro é dada e temos que calcular $N$.

$$e = z_{\alpha/2}\frac{\sigma_A}{\sqrt{N}} = 0.05$$

pelo que

$$N = \frac{z^2_{\alpha/2}\sigma^2_A}{e^2} = \frac{1.96^2 \cdot 0.45^2}{0.05^2} = 311.16 \approx 312$$

Na faculdade localizada na cidade $B$, as notas finais de curso dos candidatos ao programa de MBA acima referido seguem igualmente uma distribuição normal mas com desvio padrão igual a 0.38. Foi recolhida uma amostra aleatória de 30 candidaturas para o corrente ano, verificando-se que a média das notas finais de curso foi de 3.1.

(d) Obtenha um intervalo de confiança a 90% para a diferença entre as notas médias de curso dos candidatos ao programa de MBA nas faculdades $A$ e $B$. Interprete o resultado obtido.

**Solution:** Define-se agora $X_B$: notas finais de curso dos candidatos ao programa de MBA na faculdade $B \Longrightarrow X_B \sim \mathcal{N}\left(\mu_B, \sigma^2_B = 0.38^2\right)$
Temos que estimar $\mu_B - \mu_A$. Sabe-se que

$$Z = \frac{\overline{X}_B - \overline{X}_A - (\mu_B - \mu_A)}{\sqrt{\frac{\sigma^2_A}{N_A} + \frac{\sigma^2_B}{N_B}}} \sim \mathcal{N}(0, 1)$$

O nível de confiança é $1 - \alpha = 0.9 \iff \alpha = 0.1 \Longrightarrow z_{\alpha/2} = z_{0.05} = 1.645$. Então,

$$\begin{aligned}
[\mu_B - \mu_A]_{90\%} &= \left[\overline{X}_B - \overline{X}_A \pm z_{\alpha/2}\sqrt{\frac{\sigma^2_A}{N_A} + \frac{\sigma^2_B}{N_B}}\right] = \\
&= \left[3.1 - 2.9 \pm 1.645\sqrt{\frac{0.45^2}{25} + \frac{0.38^2}{30}}\right] = [0.013067; 0.386933]
\end{aligned}$$

Com um nível de confiança de 90% a diferença entre as notas médias finais de curso dos candidatos ao MBA nas faculdades $B$ e $A$ fica entre 0.013 e 0.3869 pontos.

(e) Sem efectuar cálculos adicionais diga qual o efeito na amplitude do intervalo acima calculado resultante do nível de confiança ser de 99% e ambas as amostras terem 35 elementos.

**Solution:** O facto de o nível de confiança subir para 99% faz com que a amplitude do intervalo aumente. O facto da dimensão de ambas as amostras aumentar faz com que a amplitude de intervalo se reduz. Assim o efeito final na amplitude de intervalo depende de qual dos dois fontes é mais forte.

2. (8 points) A Carris encomendou um estudo sobre o tempo de duração de três percursos alternativos, $A$, $B$, e $C$ que ligam dois pontos da cidade de Lisboa. Os resultados obtidos através desse estudo foram os seguintes:

   - As durações dos percursos $A$, $B$ e $C$ seguem uma distribuição normal, com igual média
   - Em 25 viagens teste efectuadas segundo o percurso $A$ obteve-se uma duração média de 30 m e uma variância corrigida de 65 $m^2$.
   - Em 25 viagens teste efectuadas segundo o percurso $B$ obteve-se uma duração média de 32 m e uma variância corrigida de 85 $m^2$.
   - Em 25 viagens teste efectuadas segundo o percurso $C$ obteve-se uma duração média de 29 m e uma variância corrigida de 54 $m^2$.

   Sabe-se que serão rejeitados os percursos que apresentem uma duração cujo desvio padrão seja superior a 7 minutos. Sabe-se ainda que, dos percursos não rejeitados, será preferido aquele que apresente uma duração com menor desvio padrão. Realizando ensaios de hipóteses com um nível de significância de 5% responda às seguintes questões:

   (a) Existe algum ou alguns percursos que sejam rejeitados logo à partida? Em caso afirmativo explique qual o erro de rejeição cometido.

   (b) Existe algum percurso que seja preferido face aos outros? Justifique.

3. Um feirante recebeu uma proposta de compra, a um preço muito atractivo, de um lote de 100 peças de tecido. O vendedor assegura-lhe que apenas 30% das peças apresentam defeito mas o feirante recebe informações segundo as quais existem 60% de peças defeituosas no lote.

   (a) Com o objectivo de realizar um teste, com 5% de significância, à hipótese apresentada pelo vendedor, de que o lote tem 30% de peças defeituosas, o feirante recolheu uma amostra (com reposição) de 36 peças, das quais 16 apresentaram defeito.

      i. Qual terá sido o resultado deste teste?

      ii. E qual o valor da potência deste teste?

   (b) Considere agora que ainda com o objectivo de testar a hipótese nula da proporção de peças defeituosas no lote ser de 30% contra a hipótese alternativa dessa mesma proporção ser de 60%, o feirante recolhe uma amostra (com reposição) de 36 peças, decidindo rejeitar o lote e não efectuar a compra caso o número de peças defeituosas na amostra seja igual ou superior a 18.

i. Quais seriam as probabilidades de cometer os erros de tipo I e II usando este critério?

ii. De que alternativas dispõe o feirante para diminuir a probabilidade de cometer o erro de tipo II?

4. (6 points) Diga se as seguintes afirmações são verdadeiras ou falsas. Justifique as suas respostas.

(a) Para uma dada amostra, o intervalo de confiança a 95% para a média é $[21, 24]$; então podemos assegurar que a probabilidade de a média estar no intervalo $[21, 24]$ é 0.95.

> **Solution:** Para uma dada amostra, o intervalo de confiança não é aleatório. Este intervalo conterá ou não a verdadeira média, mas não se pode falar de probabilidade de um número estar num intervalo numérico. Só poderá dizer-se que temos uma "confiança" de 95% da média estar no intervalo. Assim, a afirmação é **FALSA**.

(b) Para testar a hipótese nula $H_0 : \theta = 8$ versus $H_1 : \theta < 8$ decidimos usar como região de rejeição $R_C = \left\{\hat{\theta} < 7\right\}$, sendo $\hat{\theta}$ um estimador para $\theta$. Então, se o verdadeiro valor de $\theta$ é 7.5 e para a amostra dada $\hat{\theta} = 7.5$, estamos a cometer um erro tipo I.

> **Solution:** Se $\theta = 7.5 < 8$, então $H_0$ é falsa. Além disso, como $\hat{\theta} = 7.5 > 7$, rejeita-se $H_0$. Assim, não rejeitamos $H_0$ e esta é falsa e, estando a cometer um erro tipo II. Então, a afirmação é **FALSA**.

(c) Se realizamos um teste com um nível de significância de 5%, então a probabilidade de a hipótese nula ser verdadeira é 0.95.

> **Solution:** A hipótese nula é uma condição sobre o parâmetro de uma distribuição. Esta hipótese pode ou não ser verdadeira, mas como o parâmetro não é uma variável aleatória não podemos falar de probabilidade de ser verdadeira. Assim, a afirmação é **FALSA**.

(d) Seja $X_1, X_2, X_3$ uma amostra aleatória i.i.d. de uma variável aleatória $X$ com média $\mu$ e variância finita $\sigma^2$. Para estimar $\mu$ considera-se o estimador

$$\hat{\mu} = \frac{X_1 + 2X_2 + 3X_3}{6}.$$

Então $\hat{\mu}$ é um estimador não-enviesado para $\mu$.

**Solution:**

$$\begin{aligned}
\mathbb{E}(\hat{\mu}) &= \mathbb{E}\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right) = \\
&= \frac{1}{6}\left[\mathbb{E}(X_1) + 2\mathbb{E}(X_2) + 3\mathbb{E}(X_3)\right] = \\
&= \frac{1}{6}(\mu + 2\mu + 3\mu) = \mu.
\end{aligned}$$

Assim, a afirmação é **VERDADEIRA**.

(e) Para testar a hipótese nula de que as médias de três populações normais são iguais, supondo igualdade de variâncias, usa-se uma estatística que segue, sob a hipótese nula, uma distribuição $F$. Assim a região de rejeição deve ser bilateral já que a hipótese alternativa é $\mu_1 \neq \mu_2$ ou $\mu_1 \neq \mu_3$ ou $\mu_2 \neq \mu_3$.

**Solution:** Para este teste a região crítica é sempre unilateral já que sob a hipótese alternativa a estatística toma valores grandes, enquanto sob a nula toma valores pequenos. Assim, a afirmação é **FALSA**.